Title: Toward measuring biogeochemistry within the stream-groundwater interface at the network scale: an initial assessment of two spatial sampling strategies

Authors: Joseph A. Lee-Cullin[1], Jay P. Zarnetske[1], Sydney S. Ruhala[1], Stephen Plont [1]

1. Michigan State University, Department of Earth and Environmental Sciences, East Lansing, MI, USA

Running Head: Network scale stream-groundwater biogeochemistry

Keywords: surface water-groundwater, watershed, dissolved organic carbon, nitrogen, stream, biogeochemistry, scale

Abstract

It is important to understand how point measurements across spatially heterogeneous ecosystems are scaled to represent the system of interest. Stream biogeochemistry presents an illustrative example because water quality concerns within stream networks and recipient water bodies motivate heterogeneous watershed studies. Measurements of the stream water-groundwater (SW-GW) interface (i.e., the shallow subsurface of streams) are well-documented for small, point-scale sampling density measurements (i.e., cm^2-m^2 features), but poorly characterized for larger, watershed-scale sampling density measurements (i.e., km^2; stream reaches and networks). Further, sampling the SW-GW interface is more time- and labor-intensive than surface water sampling, meaning sample point selection must be made with care when attempting a network-scale analysis. In this study, we endeavor to determine which of two common spatial sampling schemes is appropriate for characterizing the biogeochemistry of the SW-GW interface across a temperate, third-order stream network, focusing on dissolved organic carbon. The first scheme, called here Local Sampling, focuses on characterization of the small-scale ($< 10\ m^2$) variability produced by the local physical and biogeochemical heterogeneity, with fewer points across the stream network. The second scheme, called here Longitudinal Sampling, has approximately the same number of measurements distributed over many more points across the stream network with less characterization of local variability. This comparison reveals that selection of a Local Sampling versus a Longitudinal Sampling scheme influences the interpretation of biogeochemical patterns at the stream network scale. Additionally, this study found an increase in observation efforts at the local scale added limited information for reach- to network-scale biogeochemical patterns, suggesting that emphasis should be placed on characterizing variability across broader spatial scales with the Longitudinal Sampling approach.

Introduction

At what spatial resolution do we make measurements and observations to characterize patterns and processes across stream networks? It is well-established in terrestrial landscape ecology that measurements made at a certain spatial sampling density (i.e., resolution or grain size) can be extrapolated to different scales of spatial extent (e.g., Schneider 1998; Wu and Li 2006). However, the best practices for extrapolating between scales are continually evolving, including many methods that have been developed to upscale or downscale observations to different resolutions (Turner and Gardner 2015). Few studies have presented best sampling practices and methods across scales for aquatic ecosystems. Streams and their interfaces with groundwater are particularly challenging for choosing the most appropriate sampling resolution due to the inherent effects of directionality in flowing water and logistical challenges of measuring surface water and groundwater parameters. To determine the ecological conditions and functioning of the stream water-groundwater (SW-GW) interface that are relevant to landscape biogeochemical budgets, watershed management, and ecosystem theories at the reach to network scales (Krause et al. 2011; Bernhardt et al. 2017), we must address how best to measure SW-GW interactions across spatial and temporal scales.

At stream network scales, the River Continuum Concept (RCC; Vannote et al. 1980) was a key step in starting to address the landscape ecology of stream networks, including the effects of directional flow through streams. The RCC postulated a gradient, moving from headwaters to higher-order streams, to explain the downstream movement and transformation of organic matter by physical and biological processes. Although aspects of the RCC are still debated (see, e.g., Creed et al. 2015; Rosi-Marshall et al. 2016), the general conceptual model of a gradient wherein the biogeochemistry of stream reaches changes systematically from upstream to downstream in networks is central to contemporary literature of stream ecosystems (e.g., Poole 2002; Thorp and Bowes 2017). The RCC is raised here not to debate its strengths and weaknesses in explaining how conditions may change through a river network, but because it did not specify what scale of measurements is needed to assess the ecological hypotheses of the RCC. Hence, there is still uncertainty in how to assess the RCC. However, in a review of stream ecology and biogeochemistry, Fisher et al. (2004) identified a broad understanding that streams are largely influenced by longitudinal (i.e., upstream/downstream) changes, and are composed of a multitude of parallel flowpaths leading to a high degree of heterogeneity. Other studies have stressed that nearby sampling points of stream chemistry were very similar, but were able to maintain a broader heterogeneous trend (Dent and Grimm 1999).

Generally, spatial biogeochemical variation in surface waters decreases with increasing stream order (Temnerud and Bishop 2005), but there is evidence in some streams that comparable variability can be found at all scales depending on the sampling density (Zimmer et al. 2013; Abbott et al. 2018). This spatial biogeochemical variability suggests that the role of study design, especially the spatial resolution of sampling, can introduce bias or confusion in our understanding of stream ecology. Furthermore, the biogeochemical variability in SW-GW interfaces across stream networks is virtually unknown and almost never documented in the literature (but see Ruhala et al. 2018). This is particularly true for assessing the structure and dynamics of the hyporheic zone (HZ), the ecotone where stream water readily interacts and exchanges properties with groundwater (Boulton et al. 1998). The HZ is a known biogeochemical control point in watersheds influencing ecosystems and water quality (McClain et al. 2003; Bernhardt et al. 2017). The primary limitation with sampling sediment pore water in the SW-GW interface is that it can be time- and labor-intensive, given that porewater must be drawn out slowly to avoid disrupting stream, hyporheic, and groundwater flow fields (i.e., typically < 5 ml min^{-1}) (e.g., Duff et al. 1998). The SW-GW interface is also known to exhibit large spatiotemporal heterogeneity in physical and biological conditions (Boano et al. 2010).

Our understanding of SW-GW interface biogeochemistry at stream network scales has been limited by a lack of understanding of how to best allocate sampling efforts in space and time. The local scale (i.e., the within-reach scale) over which SW-GW interface data are typically collected does not match the stream network scale at which many environmental problems need to be addressed (Krause et al. 2011). In fact, most SW-GW interface studies do not make direct measurements in the SW-GW interface, and instead use indirect measurements (i.e., tracer studies) that span a longitudinal scale of 10-1000 m (Ward 2016). These indirect measurements are often rife with model uncertainty and interpretation, especially for quantifying SW-GW exchange (Kelleher et al. 2013). Despite the lack of direct measurements, significant advances in process-based modeling of SW-GW processes at the stream network scale have proceeded, including the transport and fate of nutrients (Kiel and Cardenas 2014; Gomez-Velez et al. 2015). Unfortunately, there is still a paucity of data sets of the SW-GW interface at the network scale available to validate these types of models.

In thinking about a sampling scheme of the SW-GW interface across an entire stream network, one must consider the effort spent for an individual sampling point while ensuring that the limited available number of sampling points reasonably represent the entire network. Generally, there are two stream network-scale sampling schemes that appear in the literature (Figure 1): 1) high-resolution characterization of local-scale variability at few sites across the network (e.g., Zimmer et al. 2013; hereafter Local Sampling), wherein effort is focused on taking many samples at specific local-scale features in a watershed instead of fewer samples at more locations, or 2) low-resolution characterization of local-scale variability at many sites across the network, (e.g., McGuire et al. 2014; hereafter Longitudinal Sampling), wherein effort is focused on taking samples at more locations across the entire network instead of more samples at specific local-scale features. The schemes are either deliberately or arbitrarily selected to investigate properties relevant to stream network biogeochemistry. Local Sampling is often applied for investigations of specific SW-GW processes, while there are very few examples of Longitudinal Sampling studies for any type of SW-GW processes (Ward 2016). However, it is unknown whether one of these two sampling schemes is more appropriate for research questions dealing with characterization of SW-GW interface biogeochemistry at the network-scale. Our objectives in this paper are to raise awareness regarding SW-GW sampling design unknowns and to begin addressing these unknowns in our investigation of network-scale SW-GW interactions by comparing the two common sampling schemes across a stream network. Determining which scheme, Local Sampling or Longitudinal Sampling, best characterizes the overall stream network to will help advance SW-GW investigations and thus guide best sampling practices (Krause et al. 2011). To direct these main objectives, we developed the following hypotheses:

H1: A single point profile is representative of multiple point profile measurements of SW-GW interface biogeochemistry, because inter-reach variability will be greater than intra-reach variability. This hypothesis will assess whether sampling of the SW-GW interface should focus on fewer points at more sites in a network or if it is necessary to have many points to characterize each individual site, which, in turn, will guide sampling design for future SW-GW studies.

H2: Variance in SW-GW interface biogeochemistry profiles will decrease with increasing stream order, because the effects of upstream processes are integrated downstream due to directional flow. This hypothesis will help inform the development of network continuum concept in the SW-GW interface, such as the continuum concepts of the RCC.

To evaluate these objectives and hypotheses, we analyzed a spatially intensive sampling of SW-GW biogeochemistry (as compared to other SW-GW interface studies in the literature) in a stream network that spans the two study sampling schemes (Ruhala et al. 2018). Specifically, we focus on the surface water and SW-GW interface pore-water concentrations of dissolved organic carbon (DOC) in a lowland, third-order, mixed land use watershed. DOC was selected as the focus for this initial assessment because it is a fundamental control on water quality and ecosystem ecology of freshwaters due, in part, to its role in nutrient and metal cycling, ability to influence pH, effect on net carbon balances, and control of photochemistry (Aiken 2014). In addition to DOC, we include analyses for select anions, including chloride (Cl^-) and nitrate (NO_3^-) to represent nonreactive and reactive solutes, respectively (e.g., Triska et al. 1993; Barber et al. 2005; Zarnetske et al. 2011; Bernhardt et al. 2017).

Materials and Procedures

Site description – The data sets used in this study were generated by Ruhala et al. (2018) in Augusta Creek (Figure 2), which is a low gradient, third-order watershed draining 98 km^2 in southwest Michigan, USA. The watershed is composed of glacial till, and flows through a mixed-use landscape that includes wetlands, lakes, agriculture, and upland forests. The stream is primarily groundwater-fed, gaining water along much of its length, and the low overland runoff as well as abundant wetlands and lakes along its course buffer the stream discharge response to storm events (Poff et al. 1997; Hamilton et al. 2018). Stream reaches included in this study range from first- to third-order, with variable origins including lake outflows, wetland outflows, and forested headwater streams (Figure 2). Located near the W.K. Kellogg Biological Station of Michigan State University (KBS), Augusta Creek is a historically important site for freshwater biogeochemical and ecological research. For example, it was a site in the seminal RCC and Natural Flow Regime papers (Vannote et al. 1980; Poff et al. 1997), is part of the KBS Long Term Ecological Research site activities, and has an active, long-term (>50y) United States Geological Survey (USGS) gaging station (04105700).

Sampling schemes: Ruhala et al. (2018) collected data that span the Local Sampling and Longitudinal Sampling schemes, and importantly, each sampling date represented roughly the same field sampling effort (~10 field work days for 4 researchers), the same sampling techniques and equipment, and a comparable total number of SW-GW biogeochemical sample locations (n≈40). However, the team distributed these sampling points differently across the stream network, stratifying the sampling to capture most subwatersheds and all stream orders in the Augusta Creek watershed. The sampling scheme roughly corresponded to the two study scheme types, Local and Longitudinal (Figure 2).

In the data set, Local Sampling samplings characterized the local heterogeneity of a limited number of sites across the network and were carried out from 10-17 August 2015. In the Local Sampling scheme, 16 locations, stratified by stream order (first through third) were selected across the network (Figure 2). Within each location, 3 MINIPOINT porewater piezometers (Duff et al. 1998) were deployed close to each other (<3 meters apart), and hereafter the group of three samplers will be referred to as a plot (Figure 1). The MINIPOINT porewater piezometers are relatively non-invasive and allow sampling of pore water profiles from six discrete depths in the SW-GW interface (Duff et al. 1998), set between 2.5 and 20 cm as detailed in the next section and Ruhala et al. (2018). Thus, there were 18 SW-GW samples collected at each of the 16 plots for a total of 288 unique SW-GW biogeochemical sample locations from the Local Sampling approach. In Augusta Creek, most of the stream sediment is unconsolidated sandy and gravelly sediments, which is compatible with the MINIPOINT technology. However, the exact MINIPOINT porewater piezometer location at a selected site depended on the capacity to physically insert all the piezometers the specified depth into the sediment (i.e., sites with cobble or armored sediments could not be sampled).

The Longitudinal Sampling scheme represented a coarser characterization of local heterogeneity, but increased the total number of plots across the stream network and thus was meant to capture the spatial variability across the stream network. This sampling was carried out from 16-22 August 2016 during similar seasonal, stream DOC conditions, and daily discharge conditions as the Local Sampling campaign (Figure 3), though 2016 data was collected during discharge recession from a preceding high flow event. For Longitudinal Sampling, a similar field effort yielded 39 points across the network. At each location, a single MINIPOINT porewater piezometer was sampled, optimally collecting six porewater samples per point for a total of 230 unique SW-GW biogeochemical sampling locations from the Longitudinal Sampling.

Furthermore, given that we are specifically interested in the biogeochemistry with respect to DOC at larger spatial scales, we also analyzed data grouped by stream order similar to the RCC (Vannote et al. 1980). Stream order acts as a proxy for the physical hydrography of stream reaches, which in turn is fundamental to ecological patterns and processes (Harvey and Gooseff 2015). It is a simple method to discretize the network that allows for quick analysis of how an ecological variable related to DOC varies from upstream to downstream through a stream network (e.g., Creed et al. 2015). In the Local Sampling scheme there were 6 first-order, 5 second-order, and 5 third-order locations, while the Longitudinal Sampling scheme was composed of 16 first-order, 14 second-order, and 9 third-order plots. This enables an assessment of how the biogeochemistry changes with different hydrological characteristics distributed from headwaters to mainstem outlet (as addressed by H2 above).

Sample and data collection – To illustrate the procedure and effort involved in collecting SW-GW samples, here we briefly review the sampling protocol from Ruhala et al. (2018). Each MINIPOINT porewater piezometer was deployed to collect six discrete samples at 2.5, 5, 7.5, 10, 15, and 20 cm depth. The MINIPOINTs were attached to a Masterflex peristaltic pump (Cole-Parmer) using L/S Tygon tubing, and water was drawn from the SW-GW at a rate of 2.5 ml min^{-1}. They collected 80 mL of water from each depth. They used 20 mL of sample as a rinse through the filter (Whatman GF/F, 0.7 µm nominal pore size) to remove particulate matter. The remaining 60 mL was filtered through the 0.7 µm filter to remove particulates and larger microbes the placed in acid-rinsed HDPE amber bottles and stored on ice. At the end of the sampling day, 10 mL were first used to rinse through a filter (Sartorius Stedim cellulose acetate, 0.2 µm nominal pore size), then the remaining 50 mL were filtered and stored in the dark at 4°C and analyzed within 28 days. Each filtered sample was analyzed for non-purgeable organic carbon using a TOC-L total organic carbon analyzer (Shimadzu) with Pt-catalyzed oxidation at 680°C. Concentrations for Cl^- and NO_3^- were analyzed on a Dionex ICS-2100 Ion Chromatography System (ThermoScientific).

Data analysis – The Local Sampling data were divided into points, representing a single MINIPOINT with six samples at vertically distributed depths, and plots representing three

MINIPOINTs with eighteen samples, varying in depth, at a single site (Figure 4). The Longitudinal Sampling data was simply divided into points, as there was only a single MINIPOINT with six vertically distributed samples deployed at each individual site. We calculated variance for a point as the variance across the six individual depths from a MINIPOINT sampling, and variance for a plot as variance across all eighteen samples (6 depths at 3 points) from the clustered MINIPOINTs (Figure 5) as:

$$\sigma^2 = \sum \frac{(X-\mu)^2}{N} \quad (1)$$

where X is a biogeochemical concentration value at one discrete piezometer (within a MINIPOINT array for point variance and within the three MINIPOINT arrays for plot variance), μ is the mean of all concentration measurements (again, within a single MINIPOINT array for point variance and for all three MINIPOINT arrays for plot variance), and N is the number of observations (N=6 for point variance, N=18 for plot variance).

For the Local Sampling data, to assess the relative utility of a single MINIPOINT as compared to three MINIPOINTs we took the ratio of the plot variance to point variance (F), shown as:

$$F = \frac{\sigma^2_{\text{plot}}}{\sigma^2_{\text{point}}} \quad (2)$$

where σ^2 is the variance (Equation 1) and the subscripts represent the plot and points. Finally, to compare the full distributions of point and plot measurements across stream orders we used a non-parametric Wilcoxon Rank Sum Test (Wilcoxon 1945) implemented in the software R v 3.4.2 (R Core Team 2017). The Wilcoxon Rank Sum Test allows us to assess whether the distribution of samples within orders are increasing or decreasing across first, second, and third orders. This assessment is used to determine if similar patterns emerge when comparing point and plot measurements and when comparing Local Sampling to Longitudinal Sampling.

Assessment

Concentrations of DOC in the SW-GW interface were comparable between the Local Sampling and Longitudinal Sampling schemes across the network and across samplings grouped by stream order (Figure 6). Minimum and maximum SW-GW DOC concentration values for the Local Sampling were 1.50 and 15.70 mg L^{-1}, respectively, while minimum and maximum SW-GW DOC concentration values for the Longitudinal Sampling were 1.34 and 17.04 mg L^{-1}, respectively (Figure 6).

Local Sampling scheme results – Point measurements of DOC exhibited a general decrease in variance from first- to third-order (Figure 7a), where there are significant differences among first- to third-order variances ($p < 0.05$). Plot measurements of DOC also exhibited decreasing variance from first- to third-order (Figure 7b) with significant differences noted ($p < 0.05$). The DOC variance ratio, F from equation 2, ranged from 0.4 to 4.3, 0.5 to 2.3, and 0.4 to 2.4 for first-, second-, and third-order streams, respectively (Figure 8a). The corresponding median ratio values for first-, second-, and third-order streams were 1.2, 1.0, and 1.2, respectively.

Variance of NO_3^- point measurements appeared to decrease from first- to second-order and then increase from second- to third-order (Figure 7d), and were significantly different across orders ($p < 0.05$). Plot-scale variance of NO_3^- indicates a decrease from first- to second-order and a decrease from second- to third-order (Figure 7e), with first- through third-order exhibiting significant differences ($p < 0.5$). The NO_3^- variance ratio ranged from 0.4 to 40.1, 0.5 to 5.9, and 0.4 to 70.9 for first-, second-, and third-order streams, respectively (Figure 8b). The corresponding median values for first-, second-, and third-order streams were 1.3, 1.0, and 1.1, respectively.

Point measurements of Cl^- increased from first- to third-order (Figure 7g) and were significantly different ($p < 0.05$). Variances of plot measurements of Cl^- were not significantly different from first- to third-order streams (Figure 7h, $p = 0.35$). The Cl^- variance ratio ranged from 0.5 to 22.6, 0.5 to 14.7, and 0.6 to 2.4 for first-, second-, and third-order streams, respectively (Figure 8c). The corresponding median values for Cl^- in first-, second-, and third-order streams were respectively 1.0, 0.8, and 1.0.

Longitudinal Sampling scheme results – The plot variances of DOC had an apparent increase from first- to third-order streams (Figure 7c) and were significantly different among orders ($p < 0.05$). Plot variances of NO_3^- decreased from first- to third-order (Figure 7f) and were significantly different ($p < 0.05$). The plot variances of Cl^- decreased from first- to second-order (Figure 7i) and were significantly different ($p < 0.05$), but a post-hoc Dunn test (Dunn 1964) indicated that there was no significant difference between second- and third-order plot variances ($p = 0.09$).

Discussion

Our analysis of spatial heterogeneity of porewater chemistry from samples throughout the Augusta Creek network reveals several critical insights into how to best collect spatial data from the SW-GW interface at the network-scale. Further, this analysis helps demonstrate that SW-GW investigators must be cognizant of how to sample when interested in larger spatial patterns, especially when considering how stream networks remove or transform reactive biogeochemical solutes.

Guiding future sampling - The results offer an indication of how to best invest our future sampling efforts when a network-scale assessment of SW-GW interface biogeochemistry is the goal. Primarily, in Augusta Creek, we find that there is little added value in increasing characterization of the local, plot-scale spatial heterogeneity, particularly for the reactive biogeochemical components DOC and NO_3^-. The point:plot ratio in the Local Sampling scheme generally centered on a value of 1 (Figure 8) for reactive (DOC and NO_3^-) and nonreactive solutes (Cl^-), meaning that a single sampling array at a site can approximate the variance of a site as well as three separate sampling arrays at a site. In fact, new patterns of variability emerge when focusing on sampling across the stream network as opposed to more detailed local characterization (e.g., Figures 7a and 7b to 7c and Figures 7d and 7e to 7f for DOC and NO_3^-, respectively), wherein the patterns of variance moving from headwaters to downstream locations actually changes when enacting a Longitudinal Sampling scheme as compared to a Local Sampling scheme.

These results indicate that a Longitudinal Sampling scheme may be preferable to a Local Sampling scheme when investigating the biogeochemistry of the SW-GW interface at the network-scale. This finding is corroborated by two recent papers that present conceptual and reduced complexity models to understand DOC (Hotchkiss et al. 2018) and NO_3^- (Marzadri et al. 2017) processing as they move from headwater to downstream locations (i.e., from low to high order streams), including the potential differential effects of the SW-GW interface across the river network. Our assessment of the two main spatial sampling schemes for SW-GW interface and the specific results from Augusta Creek inform how future researchers can attempt to evaluate and validate these new conceptual and modeling frameworks as well as historically important frameworks such as the RCC (Vannote et al. 1980).

The variance ratios observed between the two sampling strategies suggest that point measurements are reasonably representative of plot measurements in Augusta Creek, because median values for all ratios are generally equal to unity (i.e., the ratio of the variance within a plot is close to the variance of each individual point). A Wilcoxon Rank Sum Test of the distributions of variance ratios indicates that point and plot (mean of three points) measurements are similar for most chemistry samples, with the exception of DOC and NO_3^- in third-order reaches. However, the median values of plot to point ratios in third-order streams are still relatively close to unity (1.23 for DOC, 0.99 for Cl^-, and 1.12 for NO_3^-). Therefore, for this stream network under summer baseflow conditions, the results suggest that the SW-GW interface biogeochemistry of first- and second-order streams can be characterized with less focus on the local intra-site heterogeneity, which allows more focus on the inter-reach heterogeneity. In other words, more valuable data about network-scale SW-GW biogeochemical conditions can be collected using the Longitudinal Sampling scheme as compared to the Local Sampling scheme.

The observed reduction in variances of porewater concentrations moving downstream was dependent upon the biogeochemical species of interest. In the Local Sampling campaign, DOC variance at different sampling densities generally decreased moving from first- to third-order streams (Figure 7a and b). Conversely, Cl^- variance in Local Sampling increased from first- to third-order streams for both sampling densities (Figure 7g and h). NO_3^- variance exhibited an inconsistent trend in Local Sampling, wherein it increased from first- to second-order, then decreased from second- to third-order (Figure 7d and e). In Longitudinal Sampling the NO_3^- variance generally decreased with increasing stream-order (Figure 7c). The reduction in DOC variance with increasing stream order reflects the accumulation and mixing of all upstream inputs (Abbott et al. 2018). Synthesis studies of DOC across stream networks indicate that, indeed, the variability of DOC typically decreases with an increase in disconnection from terrestrial sources (e.g., Creed et al. 2015).

Most stream networks have the majority of total stream length in first- and second-order streams (e.g., first-order = 52% and second-order = 25%, Downing et al. 2012), so the finding that the low-order streams in Augusta Creek can be characterized with less focus on intra-site heterogeneity means that more low-order locations should be sampled (i.e., the Longitudinal Sampling scheme), rather than investing efforts in plot replication at each location. Historically, SW-GW interface research has disproportionately focused on second-, third-, and fourth-order streams (Ward 2016), so more effort should be directed to first and >fifth-order stream SW-GW interfaces in networks if we are to better represent SW-GW conditions in future network scale biogeochemical studies and models. Headwaters are demonstrably important in terms of the contribution of biogeochemical processes to downstream nutrient export (Alexander et al. 2007; Boano et al. 2014). Further, smaller networks tend to display the highest variability in water quality (Wolock et al. 1997; Temnerud and Bishop 2005; Abbott et al. 2018).

Different sampling resolution concerns - The Longitudinal Sampling campaign, with low local characterization in favor of higher longitudinal spatial resolution across the stream network, can potentially result in an entirely different interpretation of SW-GW conditions and DOC stream processing. DOC and Cl^- trends across orders were the opposite as compared to the trends observed in the Local Sampling scheme. Here, DOC variance is generally increasing, while Cl^- is generally decreasing moving from upstream to downstream (Figure 7c, f). While Cl^- fits our hypothesis (H2), DOC does not support it. This is an important revelation given that the same stream system was sampled under similar weather and hydrologic conditions (albeit in a different year), but changing the spatial SW-GW sampling scheme yielded a completely different apparent pattern across the network. These fundamental differences moving from headwaters to downstream locations have raised concerns particularly for empirical and mechanistic modeling. If data input into a model has a different pattern of variance depending on the sampling scheme, then the results of those models and conclusions that can be drawn from them will be entirely different from one scheme to the next.

Though studies comparing biogeochemistry at different scales are generally absent from the literature for the SW-GW interface, several researchers have identified the importance of scale in studies of SW-GW interface processes. The concerns of how sampling resolution will impact attempts to interpret or model the biogeochemical function of SW-GW interactions is more important than ever now that data users, including modelers, managers, and decision makers, are often thinking at river network scales (Krause et al., 2011). This will lead to an increase in demand for river network scale SW-GW biogeochemical data, and those seeking to collect that data must grapple with sampling effort and how resolution of sampling can impact the various data users. While SW-GW biogeochemical investigations at the network scale are limited, there are complementary ecological studies that offer further guidance. Ecological researchers have long known that different processes are scale-dependent and the scale at which one measures should answer the question being asked (e.g., Allen and Starr 1982; Delcourt et al. 1982). River corridor investigators addressing different research questions have observed spatial-resolution and extent dependent patterns, for example, small-scale biotic diversity as compared to larger-scale diversity in the SW-GW interface (see review paper by Vinson and Hawkins 1998) or comparing the riparian subsurface flow paths in small vs. large scales (see Dahl et al. 2007). Because it is important to understand all ecological processes at a variety of scales, the present study endeavored to assess how to best measure at an unprecedented network-scale in the SW-GW interface. This study helps raise some potential concerns about sampling schemes and their impact on understanding the SW-GW interface across spatial scales, and therefore should help guide future research interested in collecting and using data to compare processes across spatial scales. It also underscores that researchers cannot ignore that they must carefully consider what spatial sampling scheme may be best for the SW-GW question being asked.

A need for more assessment of sampling schemes - This study has a couple notable limitations that must be acknowledged in assessing the key differences between a Local Sampling and Longitudinal Sampling schemes. First and foremost, both studies from the Ruhala et al. (2018) data sets are snapshots in time. While they sampled at approximately the same time of year and season for Local Sampling and Longitudinal Sampling schemes, they are not capturing any of the potential sub-annual temporal dynamics of the biogeochemistry in the SW-GW interface. In SW-GW interfaces, the biogeochemistry is typically highly variable in relation to seasonal variation in nutrients, organic matter quantity and quality, and flow conditions. For example, Lambert et al. (2013) found that low aromaticity DOC accumulated in the HZ in the summer and was replaced in the wet season by more aromatic DOC, updating earlier research that had concluded that seasonal removal of DOC was relatively stable (Findlay and Sobczak 1996). Others have found that NO_3^- removal in the HZ is highly variable and dependent upon the distribution of precipitation across different seasons, as precipitation controls both productivity and routing of water through the HZ (Rahimi et al. 2015). In part, the biogeochemical variability found in this study may be due to flow variation between the two Ruhala et al. (2018) sampling periods as they observed similar biogeochemical conditions in the surface and groundwaters between sampling periods. Additionally, some variability could be due to the imprecise site selection from one year to the next, where the Longitudinal Sampling samples, while selected to overlap with the Local Sampling sites, were not taken at the exact same locations. However, given that Ruhala et al. (2018) attempted to collect at approximately the same locations both years and the results from the Local Sampling indicating that variability is fairly well-characterized by a single MINIPOINT at a location as compared to three MINIPOINTs at the same location, we expect that the variability captured in the Longitudinal Sampling should reflect the specific site from year to year.

This difference in flow conditions raises a second notable limitation to this study in that it is a comparison between two separate years. While Ruhala et al. (2018) attempted to carry out the study at similar times and seasonal conditions in each year, the hydrologic conditions were not identical, nor will they ever be in most stream systems between different sampling events. In many stream systems, shifts from high to low or low to high flow conditions can weaken or even reverse SW-GW exchange patterns (e.g., Wroblicky et al. 1998; Boano et al. 2010) as well as change the quantity and quality of solutes delivered to the SW-GW interface, such as DOC (e.g., Byrne et al. 2014; Fasching et al. 2015). Many of the limitations listed above are allayed due to the well documented hydrologic stability of Augusta Creek (e.g., Poff et al. 1997). Given that the majority of Augusta Creek stream water arrives in the channel through groundwater flowpaths (Hamilton et al. 2018), the surface water flow fluctuations and impacts on the SW-GW exchange patterns are buffered and minimized. This is to say, many of the variable flow and storm response effects commonly seen in the SW-GW interface of other streams are attenuated by the consistent groundwater inputs in this particular stream system and do not seem to shift the overall biogeochemical conditions of the stream (Figure 8). Consequently, despite these potential limitations with the data, we think that the comparison of the Local Sampling and Longitudinal Sampling data sets is useful and informative for assessing how the two sampling schemes yield different information, especially given that there is a paucity of network-scale SW-GW biogeochemical assessments available.

In many cases available data do not exist or, in the case of Ruhala et al. (2018), are not ideal for comparing Local Sampling and Longitudinal Sampling schemes. Therefore, in the future, if there were sufficient people and equipment to conduct simultaneous sampling using both Local Sampling and Longitudinal Sampling schemes, it would make for a more robust assessment of the strengths and weaknesses of each sampling scheme as well as tests of our hypotheses. Still, the present study results suggest that this larger investment in testing each study scheme is likely warranted, because it may illustrate that different network-scale patterns of the SW-GW interface biogeochemistry appear depending upon where you sample in the stream network, and inform how researchers and water quality managers can expand methods to conduct SW-GW studies at larger scales compatible with current watershed management plans and models (Hester and Gooseff 2010; Krause et al. 2011; Harvey and Gooseff 2015).

Comments and recommendations

Based on the findings in this study, we recommend an increased focus on spatial sampling schemes in SW-GW studies. We also found evidence in the study watershed that longitudinal sampling of the SW-GW interface in favor of characterizing local heterogeneity when one is interested in characterizing the SW-GW interface across a network. We must find the most efficient means of sampling, because SW-GW sampling is highly demanding of both labor and costs. From our initial assessment here, we determined that there was not much added value (i.e., detection of biogeochemical variability) with an increased effort in the characterization of local plot-scale heterogeneity (Local Sampling). There were, however, new biogeochemical patterns revealed in the watershed as the sampling scheme shifted to increase the number of plots sampled in longitudinal directions (Longitudinal Sampling), because it allowed the same sampling effort to be distributed across more of the stream network.

Overall, there is a need to investigate what the best practices are for collecting SW-GW interface data at watershed scales. Without data from the SW-GW interface at the scales of watersheds and across river networks, if may not be possible to assess and upscale the ecological function that SW-GW interfaces play in network-scale processes, such as nutrient budgets and water quality management (Harvey and Gooseff 2015; Abbott et al. 2016). As highlighted here, a clear, current limitation to assessing the role of SW-GW interfaces in river corridors is the absence of studies of the SW-GW interface attempted at a watershed scale. Hence, a possible way forward is to collect more SW-GW interface data sets at the stream network-scale from different study regions and from a particular stream network across different seasons. Overcoming this data gap will permit future researchers to evaluate if our findings from the Augusta Creek data set are robust in terms of sampling strategy suggestions and, importantly, facilitate assessments of current sampling effort utility and inspire new sampling strategies.

References and Citations

Abbott, B. W., V. Baranov, C. Mendoza-Lera, and others. 2016. Using multi-tracer inference to move beyond single-catchment ecohydrology. Earth-Science Rev. 160: 19–42. doi:10.1016/J.EARSCIREV.2016.06.014

Abbott, B. W., G. Gruau, J. P. Zarnetske, and others. 2018. Unexpected spatial stability of water chemistry in headwater stream networks J. Grover [ed.]. Ecol. Lett. 21: 296–308. doi:10.1111/ele.12897

Aiken, G. R. 2014. Dissolved organic matter in aquatic systems, p. 205–220. In S. Ahuja [ed.], Comprehensive Water Quality and Purification. Elsevier Inc.

Alexander, R. B., E. W. Boyer, R. A. Smith, G. E. Schwarz, and R. B. Moore. 2007. The Role of Headwater Streams in Downstream Water Quality. J. Am. Water Resour. Assoc. 43: 41–59. doi:10.1111/j.1752-1688.2007.00005.x

Allen, T. F. H., and T. B. Starr. 1982. Hierarchy : perspectives for ecological complexity, T. Allen and T.B. Starr [eds.]. University of Chicago PressChicago, IL.

Barber, L. B., S. F. Murphy, P. L. Verplanck, M. W. Sandstrom, H. E. Taylor, and E. T. Furlong‡. 2005. Chemical Loading into Surface Water along a Hydrological, Biogeochemical, and Land Use Gradient: A Holistic Watershed Approach.doi:10.1021/ES051270Q

Bernhardt, E. S., J. R. Blaszczak, C. D. Ficken, M. L. Fork, K. E. Kaiser, and E. C. Seybold. 2017. Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept. Ecosystems 20: 665–682. doi:10.1007/s10021-016-0103-y

Boano, F., a. Demaria, R. Revelli, and L. Ridolfi. 2010. Biogeochemical zonation due to intrameander hyporheic flow. Water Resour. Res. 46: n/a-n/a. doi:10.1029/2008WR007583

Boano, F., J. W. Harvey, A. Marion, A. I. Packman, R. Revelli, L. Ridolfi, and A. Wörman. 2014. Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. Rev. Geophys. 52: 603–679. doi:10.1002/2012RG000417

Boulton, A. J., S. Findlay, P. Marmonier, E. H. Stanley, and H. M. Valett. 1998. The functional significance of the hyporheic zone in streams and rivers. Annu. Rev. Ecol. Syst. 29: 59–81. doi:10.1146/annurev.ecolsys.29.1.59

Byrne, P., A. Binley, A. Heathwaite, and others. 2014. Control of river stage on the reactive chemistry of the hyporheic zone. Hydrol. Process. 28: 4766–4779. doi:10.1002/hyp.9981

Creed, I. F., D. M. McKnight, B. A. Pellerin, and others. 2015. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. Can. J. Fish. Aquat. Sci. 150416143426006. doi:10.1139/cjfas-2014-0400

Dahl, M., B. Nilsson, J. H. Langhoff, and J. C. Refsgaard. 2007. Review of classification systems and new multi-scale typology of groundwater–surface water interaction. J. Hydrol. 344: 1–16. doi:10.1016/J.JHYDROL.2007.06.027

Delcourt, H. R., P. A. Delcourt, and T. Webb. 1982. Dynamic plant ecology: the spectrum of vegetational change in space and time. Quat. Sci. Rev. 1: 153–175. doi:10.1016/0277-3791(82)90008-7

Dent, C. L., and N. B. Grimm. 1999. Spatial heterogeneity of stream water nutrient concentrations over successional time. Ecology 80: 2283–2298. doi:10.1890/0012-9658(1999)080[2283:SHOSWN]2.0.CO;2

Downing, J. A., J. J. Cole, C. M. Duarte, and others. 2012. Global abundance and size distribution of streams and rivers. Inl. Waters 2: 229–236. doi:10.5268/IW-2.4.502

Duff, J. H., F. Murphy, C. C. Fuller, F. J. Triska, J. W. Harvey, and A. P. Jackman. 1998. A mini drivepoint sampler for measuring pore-water solute concentrations in the hyporheic zone of sand-bottom streams. Limnol. Oceanogr. 43: 1378–1383. doi:10.4319/lo.1998.43.6.1378

Dunn, O. J. 1964. Multiple Comparisons Using Rank Sums. Technometrics 6: 241–252. doi:10.1080/00401706.1964.10490181

Fasching, C., A. J. Ulseth, J. Schelker, G. Steniczka, and T. J. Battin. 2015. Hydrology controls dissolved organic matter export and composition in an Alpine stream and its hyporheic zone. Limnol. Oceanogr. n/a-n/a. doi:10.1002/lno.10232

Findlay, S., and W. V Sobczak. 1996. Variability in removal of dissolved organic carbon in hyporheic sediments. J. North Am. Benthol. Soc. 15: 35–41.

Fisher, S. G., R. A. Sponseller, and J. B. Heffernan. 2004. Horizons in stream biogeochemistry: flowpaths to progress. Ecology 85: 2369–2379. doi:10.1890/03-0244

Gomez-Velez, J. D., J. W. Harvey, M. B. Cardenas, and B. Kiel. 2015. Denitrification in the Mississippi River network controlled by flow through river bedforms. Nat. Geosci. 8: 941–945. doi:10.1038/ngeo2567

Hamilton, S. K., M. Z. Hussain, C. Lowrie, B. Basso, and G. P. Robertson. 2018. Evapotranspiration is resilient in the face of land cover and climate change in a humid temperate catchment. Hydrol. Process. 32: 655–663. doi:10.1002/hyp.11447

Harvey, J., and M. Gooseff. 2015. River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resour. Res. 51: 6893–6922. doi:10.1002/2015WR017617

Hester, E. T., and M. N. Gooseff. 2010. Moving Beyond the Banks: Hyporheic Restoration Is Fundamental to Restoring Ecological Services and Functions of Streams. Environ. Sci. Technol. 44: 1521–1525. doi:10.1021/es902988n

Hotchkiss, E. R., S. Sadro, and P. C. Hanson. 2018. Toward a more integrative perspective on carbon metabolism across lentic and lotic inland waters. Limnol. Oceanogr. Lett. 3: 57–63. doi:10.1002/lol2.10081

Kelleher, C., T. Wagener, B. McGlynn, A. S. Ward, M. N. Gooseff, and R. A. Payn. 2013. Identifiability of transient storage model parameters along a mountain stream. Water Resour. Res. 49: 5290–5306. doi:10.1002/wrcr.20413

Kiel, B. A., and M. B. Cardenas. 2014. Lateral hyporheic exchange throughout the Mississippi

River network. Nat. Geosci. 7: 413–417. doi:10.1038/ngeo2157

Krause, S., D. M. Hannah, J. H. Fleckenstein, and others. 2011. Inter-disciplinary perspectives on processes in the hyporheic zone. Ecohydrology 4: 481–499. doi:10.1002/eco.176

Lambert, T., A.-C. Pierson-Wickmann, G. Gruau, A. Jaffrezic, P. Petitjean, J.-N. Thibault, and L. Jeanneau. 2013. Hydrologically driven seasonal changes in the sources and production mechanisms of dissolved organic carbon in a small lowland catchment. Water Resour. Res. 49: 5792–5803. doi:10.1002/wrcr.20466

Marzadri, A., M. M. Dee, D. Tonina, A. Bellin, and J. L. Tank. 2017. Role of surface and subsurface processes in scaling N2O emissions along riverine networks. Proc. Natl. Acad. Sci. U. S. A. 114: 4330–4335. doi:10.1073/pnas.1617454114

McClain, M. E., E. W. Boyer, C. L. Dent, and others. 2003. Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems. Ecosystems 6: 301–312. doi:10.1007/s10021-003-0161-9

McGuire, K. J., C. E. Torgersen, G. E. Likens, D. C. Buso, W. H. Lowe, and S. W. Bailey. 2014. Network analysis reveals multiscale controls on streamwater chemistry. Proc. Natl. Acad. Sci. U. S. A. 111: 7030–5. doi:10.1073/pnas.1404820111

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg. 1997. The Natural Flow Regime. Bioscience 47: 769–784. doi:10.2307/1313099

Poole, G. C. 2002. Fluvial landscape ecology: addressing uniqueness within the river discontinuum. Freshw. Biol. 47: 641–660. doi:10.1046/j.1365-2427.2002.00922.x

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-Project.org

Rahimi, M., H. I. Essaid, and J. T. Wilson. 2015. The role of dynamic surface water-groundwater exchange on streambed denitrification in a first-order, low-relief agricultural watershed. Water Resour. Res. 51: n/a-n/a. doi:10.1002/2014WR016739

Rosi-Marshall, E. J., K. L. Vallis, C. V. Baxter, and J. M. Davis. 2016. Retesting a prediction of the River Continuum Concept: autochthonous versus allochthonous resources in the diets of invertebrates. Freshw. Sci. 35: 534–543. doi:10.1086/686302

Ruhala, S. S., J. P. Zarnetske, D. T. Long, J. A. Lee-Cullin, S. Plont, and E. R. Wiewiora. 2018. Exploring dissolved organic carbon cycling at the stream–groundwater interface across a third-order, lowland stream network. Biogeochemistry 137: 105–126. doi:10.1007/s10533-017-0404-z

Schneider, D. C. 1998. Applied scaling theory, p. 253–270. In Ecological scale: theory and applications.

Temnerud, J., and K. Bishop. 2005. Spatial Variation of Streamwater Chemistry in Two Swedish Boreal Catchments: Implications for Environmental Assessment.doi:10.1021/ES040045Q

Thorp, J. H., and R. E. Bowes. 2017. Carbon Sources in Riverine Food Webs: New Evidence from Amino Acid Isotope Techniques. Ecosystems 20: 1029–1041. doi:10.1007/s10021-016-0091-y

Triska, F. J., J. H. Duff, and R. J. Avanzino. 1993. The role of water exchange between a stream channel and its hyporheic zone in nitrogen cycling at the terrestrial-aquatic interface. Hydrobiologia 251: 167–184. doi:10.1007/BF00007177

Turner, M. G., and R. H. Gardner. 2015. Introduction to Landscape Ecology and Scale, p. 1–32. In Landscape Ecology in Theory and Practice. Springer New York.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The River Continuum Concept. Can. J. Fish. Aquat. Sci. 37: 130–137. doi:10.1139/f80-017

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of Stream Insects: Variation at Local, Basin, and Regional Scales. Annu. Rev. Entomol. 43: 271–293. doi:10.1146/annurev.ento.43.1.271

Ward, A. S. 2016. The evolution and state of interdisciplinary hyporheic research. Wiley Interdiscip. Rev. Water 3: 83–103. doi:10.1002/wat2.1120

Wilcoxon, F. 1945. Individual Comparisons by Ranking Methods. Biometrics Bull. 1: 80–83.

Wolock, D. M., J. Fan, and G. B. Lawrence. 1997. Effects of basin size on low-flow stream chemistry and subsurface contact time in the Neversink River Watershed, New York. Hydrol. Process 11: 1273–1286.

Wroblicky, G. J., M. E. Campana, H. M. Valett, and C. N. Dahm. 1998. Seasonal variation in surface-subsurface water exchange and lateral hyporheic area of two stream-aquifer systems. Water Resour. Res. 34: 317–328. doi:10.1029/97WR03285

Wu, J., and H. Li. 2006. Concepts of scale and scaling, p. 3–15. In Scaling and Uncertainty Analysis in Ecology. Springer Netherlands.

Zarnetske, J. P., R. Haggerty, S. M. Wondzell, and M. A. Baker. 2011. Dynamics of nitrate production and removal as a function of residence time in the hyporheic zone. J. Geophys. Res. 116: G01025. doi:10.1029/2010JG001356

Zimmer, M. A., S. W. Bailey, K. J. McGuire, and T. D. Bullen. 2013. Fine scale variations of surface water chemistry in an ephemeral to perennial drainage network. Hydrol. Process. 27: 3438–3451. doi:10.1002/hyp.9449

Acknowledgments

Thank you to Michigan State University Department of Earth and Environmental Sciences, Michigan State University Environmental Science and Policy Program, Michigan State University Kellogg Biological Station Long Term Ecological Research site, and the National Science Foundation Award EAR-1446328 for research support. Thank you to S. Hamilton for helpful comments in developing sampling methodology and comments on this manuscript. Thank you to A. Shogren for proofreading the final draft. Thank you to T. Bigg, E. Wiewiora, and T. Hampton for assistance in collecting samples and to the journal editor and reviewers whose constructive comments improved the paper.

Figure Legends

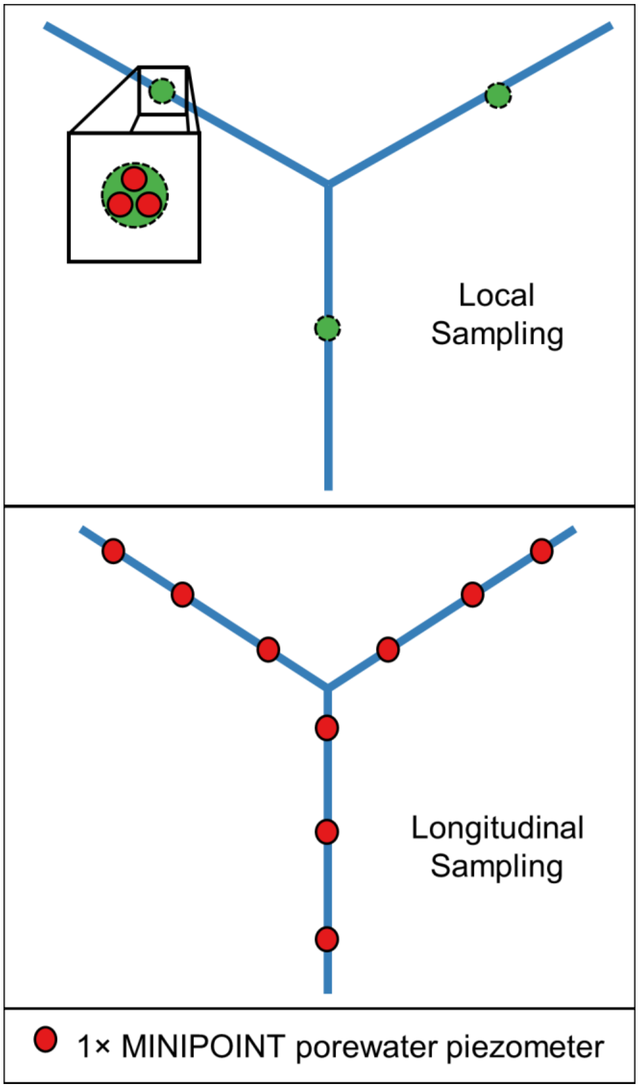


Figure 1 – Simplified plan view of stream network reaches illustrating the main conceptual differences for Local Sampling (L) and Longitudinal Sampling (R) sampling schemes. Local Sampling represents high characterization of local heterogeneity with low characterization of longitudinal heterogeneity, while Longitudinal Sampling has low characterization of local heterogeneity and high characterization of longitudinal heterogeneity. Note that each MINIPOINT sample location includes up to six depths of porewater samples in the present study.

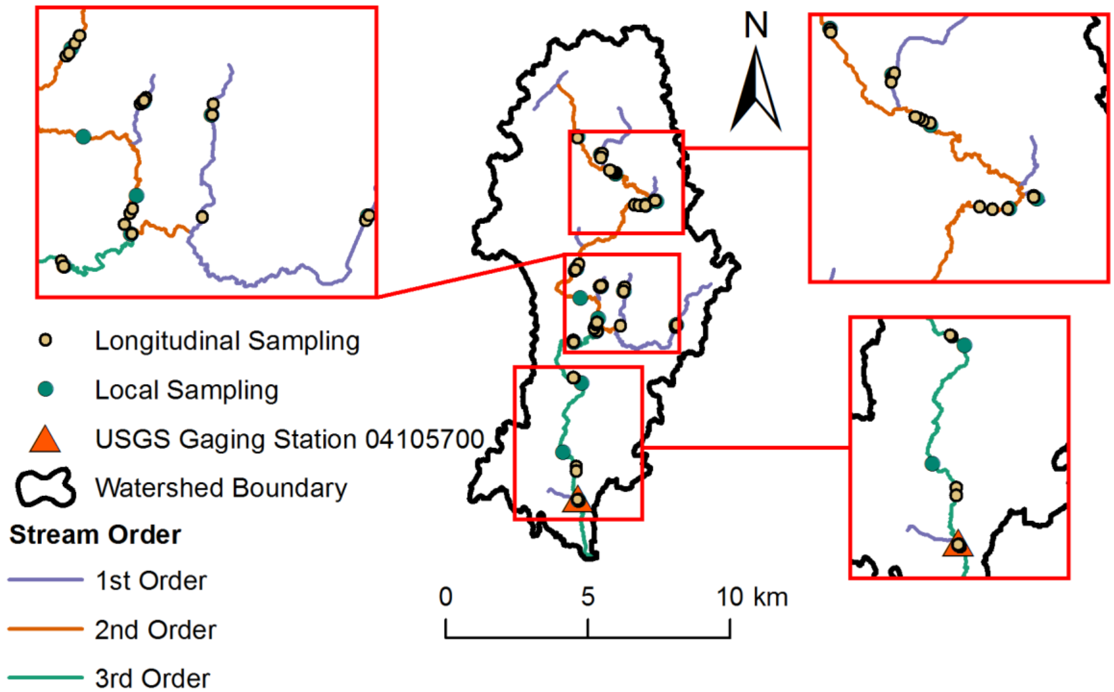


Figure 2 – Map illustrating sediment porewater sampling locations for the Local Sampling and Longitudinal Sampling campaigns, where the large, green circle symbols are the Local Sampling scheme locations, and small, yellow circle symbols are the Longitudinal Sampling scheme locations. Stream orders are identified by color, where first order streams are purple, second order streams are orange, and third order streams are green.

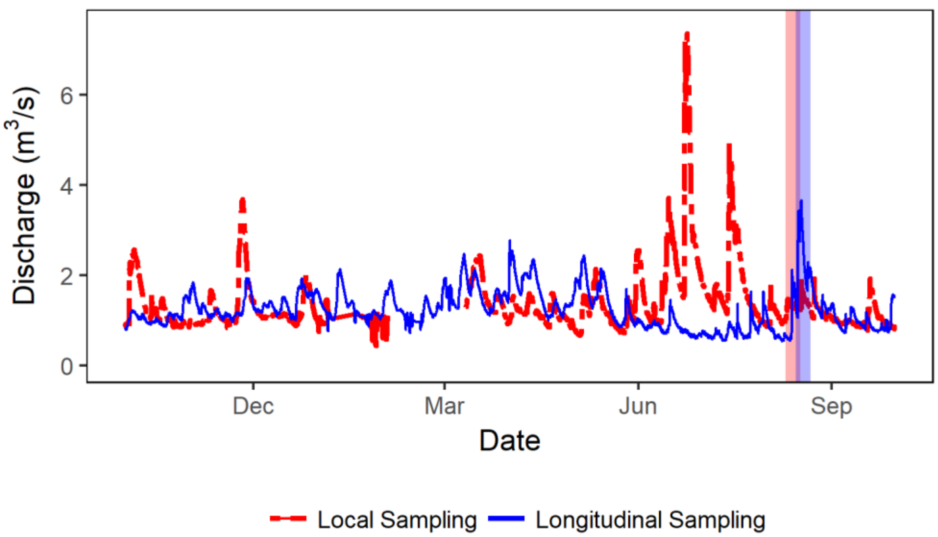


Figure 3 – Discharge conditions at the downstream USGS gaging station on Augusta Creek (04105700) for water years 2015 (red) and 2016 (blue), with shading corresponding to the Local Sampling (red) and Longitudinal Sampling (blue) efforts described in this study.

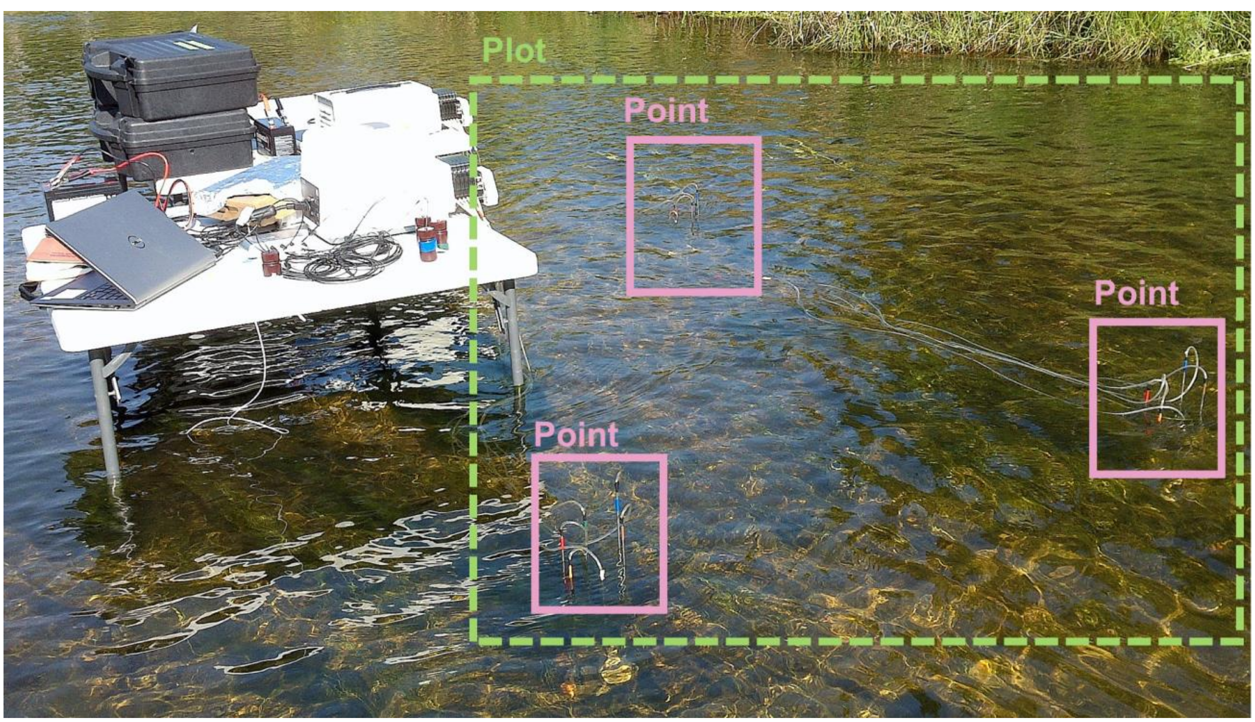


Figure 4 – Field example of the division between "points" and "plots." A point representing a single MINIPOINT array at a site and a plot representing all three MINIPOINT arrays at a site under Local Sampling scheme, whereas there would only be one MINIPOINT array point in a plot under Longitudinal Sampling scheme.

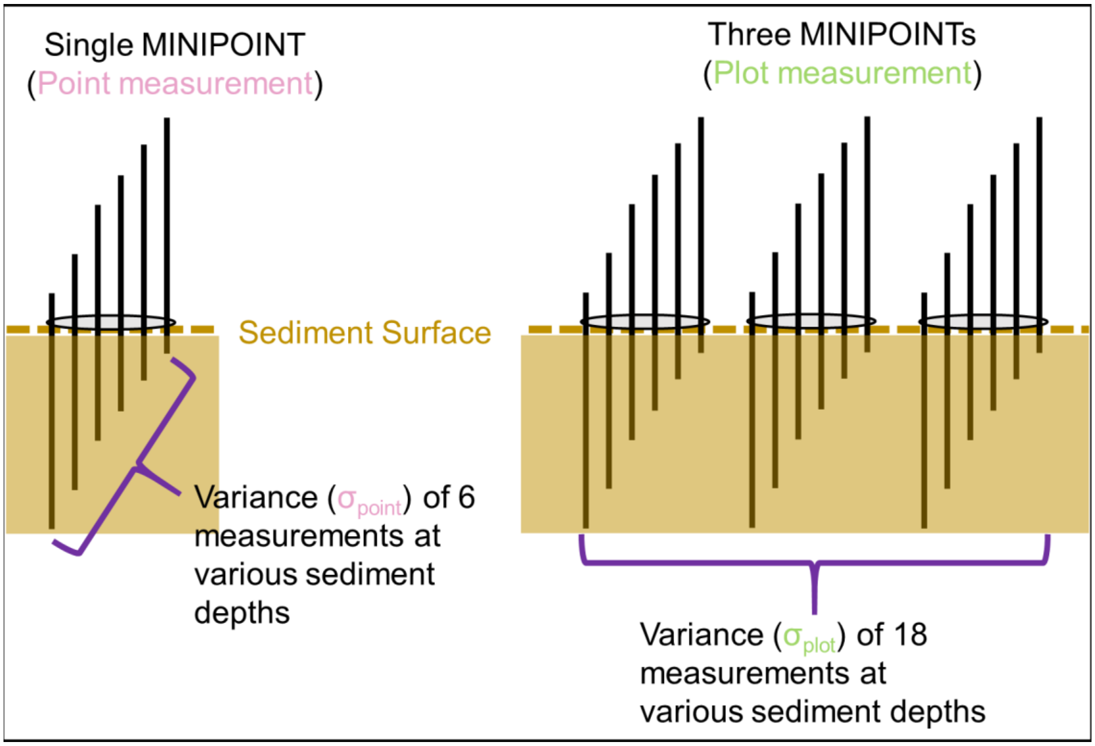


Figure 5 – Illustration of the distinction between point variance (on left) and plot variance (on right) in this study. Point variance represents the variance of 6 discrete depths of a single MINIPOINT, while plot variance represents the variance between all 18 measurements of the three MINIPOINTs at a site.

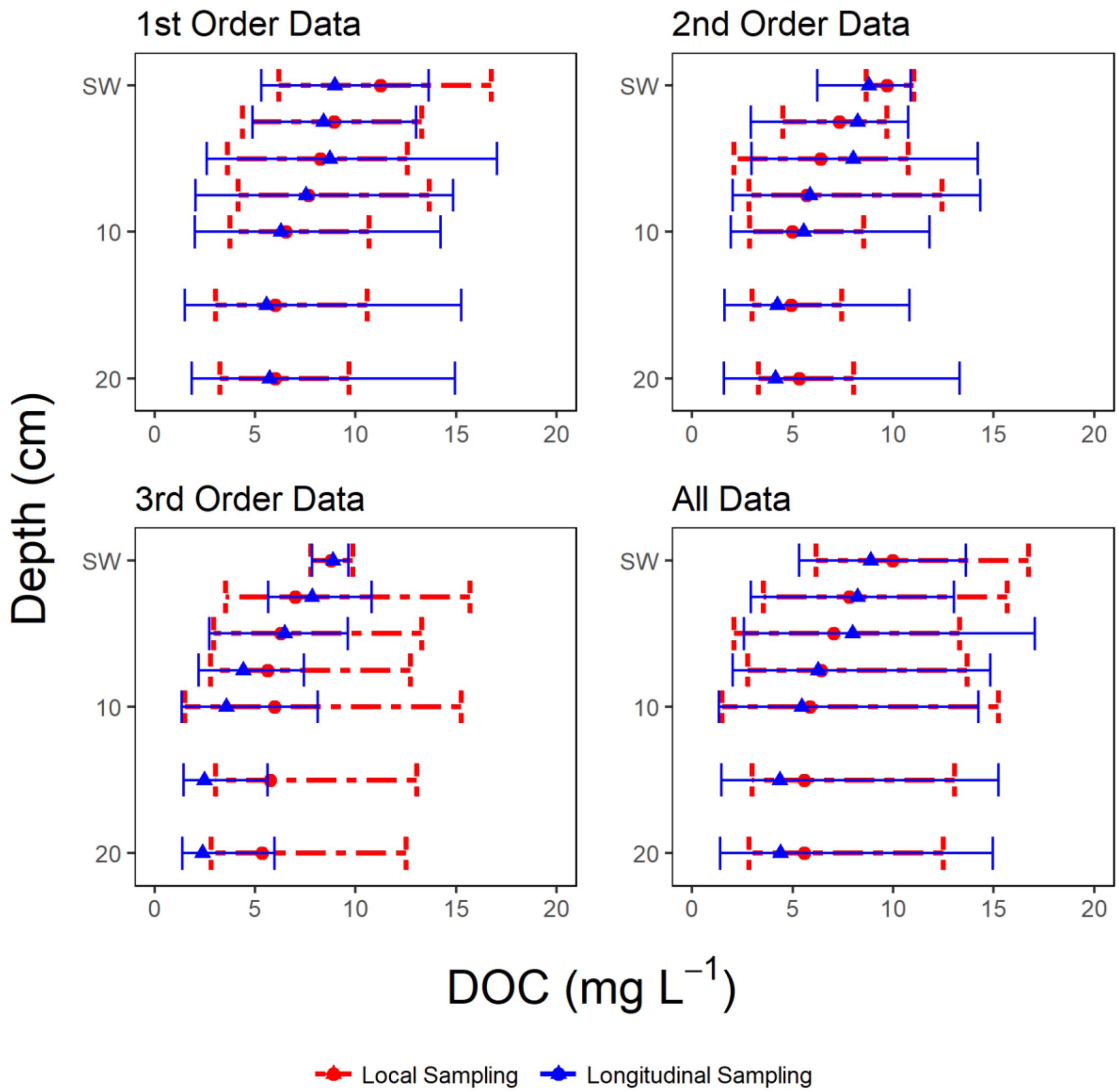
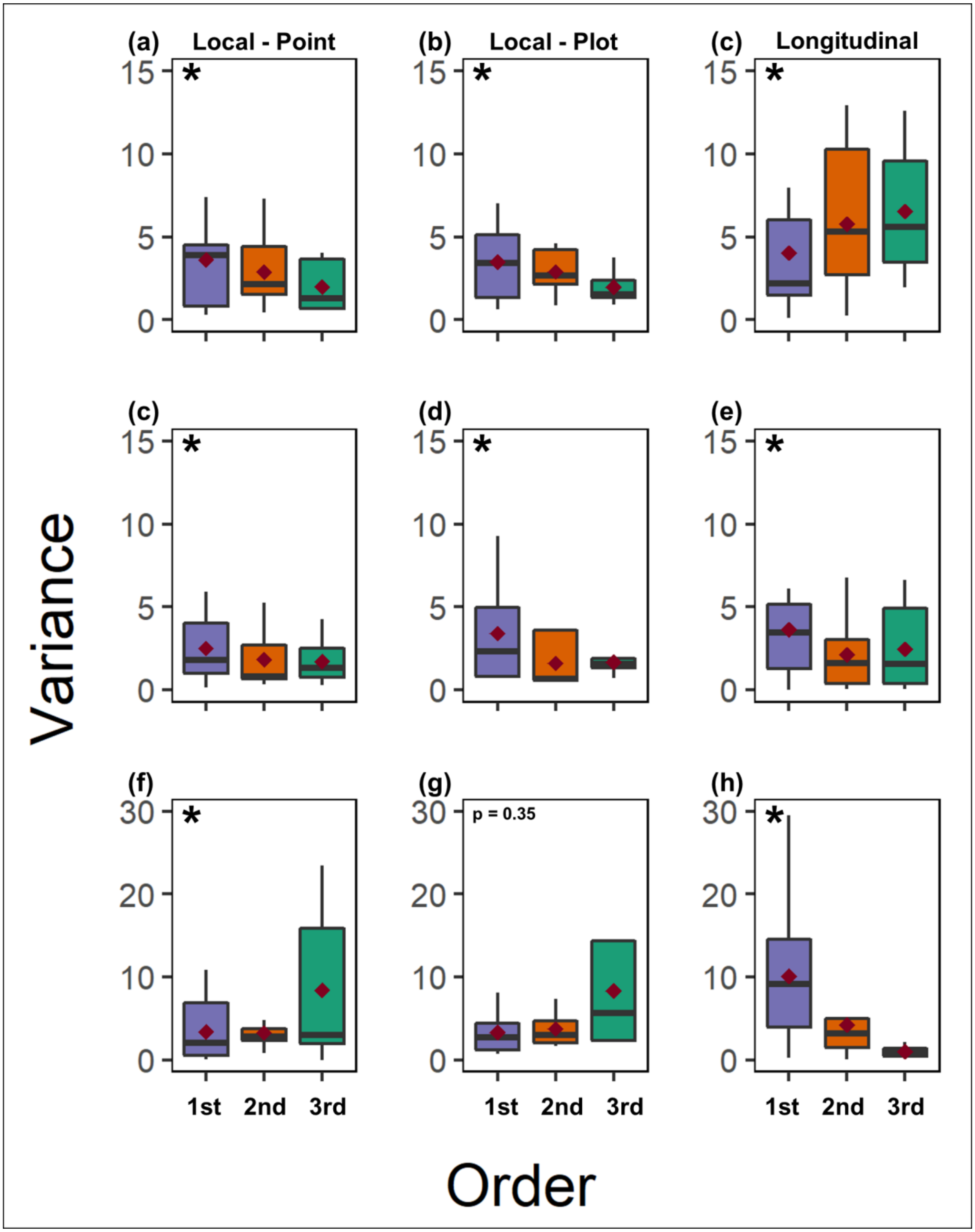


Figure 6 – Point and whisker plots representing the mean (points) and range (whiskers) of observed porewater dissolved organic carbon (DOC) concentrations in Augusta Creek (all depths included) for both Local Sampling and Longitudinal Sampling schemes across the all of the network and grouped by stream order.

652 

653 Figure 7 – Box and whisker plots illustrating the distribution of variance for Local Sampling

654 (i.e., high local characterization) and Longitudinal Sampling (i.e., low local characterization, but

greater longitudinal characterization) for measurements of dissolved organic carbon (DOC; a-c), NO_3^- (d-f), and Cl^- (g-i) at points (a single MINIPOINT at a site) and plots (three MINIPOINTs at a site) across first, second, and third-order reaches of the Augusta Creek system. Distributions of the same sampling type are all significantly different per a Wilcoxon Rank Sum Test ($p < 0.05$), as noted with an * or otherwise stated with the specific p-value.

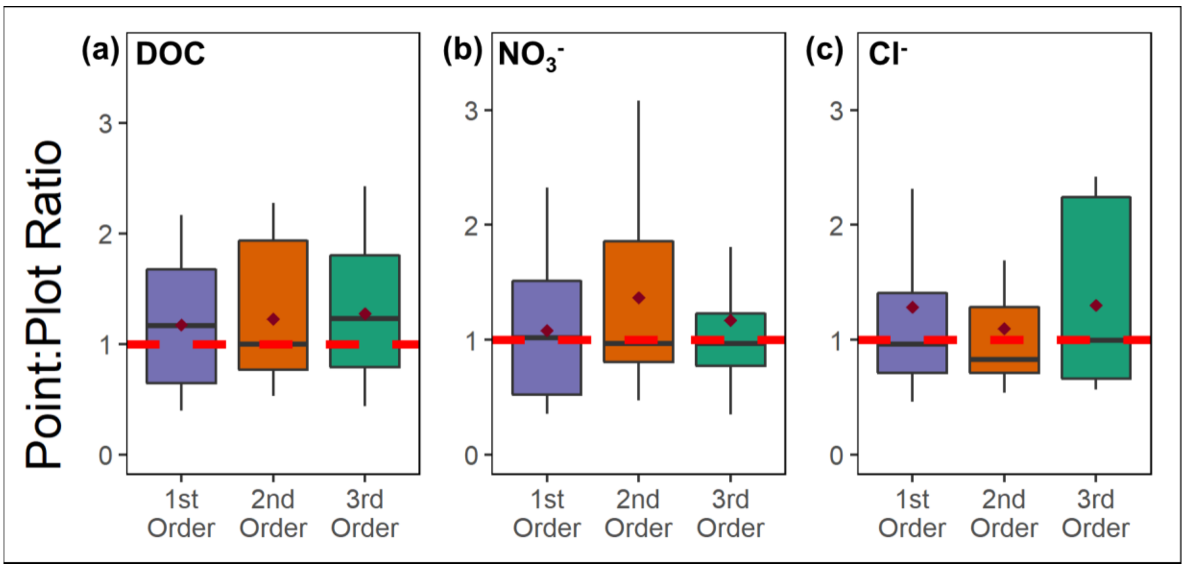


Figure 8 – Point (single MINIPOINT) to plot (three MINIPOINTs) variance ratios across stream orders during the Local Sampling (high local characterization) sampling campaign. The box and whiskers represent the quartiles at each stream order for the Local Sampling scheme, with the solid line indicating median values. The red diamonds represent mean values. Ratio values less than 1 indicate point variability is greater than plot variability, values greater than 1 indicate that point variability is less than plot variability, and values equal to 1 indicate point variability is equal to plot variability. The red dashed line represents a value of 1.